In [14]:
```python
#Aim : To perform Data Processing,Data Cleaning,Missing Value Treatemen
```

In [2]:
```python
# Name: Pragati Pramod Bindod
# Roll no. : 15
# Section : A
```

In [3]:
```python
import pandas as pd
```

In [4]:
```python
data=pd.read_csv("C:\\Users\\PRAGATI BINDOD\\Downloads\\archive (1).zip
```

In [5]: ▶ | data

Out[5]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket
0	1	0	3	Braund, Mr. Owen Harris	male	22.0	1	0	A/5 21171
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599
2	3	1	3	Heikkinen, Miss. Laina	female	26.0	0	0	STON/O2. 3101282
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803
4	5	0	3	Allen, Mr. William Henry	male	35.0	0	0	373450
...	...	...	...	...	...	...	...	...	...
886	887	0	2	Montvila, Rev. Juozas	male	27.0	0	0	211536
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053
888	889	0	3	Johnston, Miss. Catherine Helen "Carrie"	female	NaN	1	2	W./C. 6607
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369
890	891	0	3	Dooley, Mr. Patrick	male	32.0	0	0	370376

891 rows × 12 columns

In [15]: ▶| `data.head()`

Out[15]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	F
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71.2
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53.1
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51.8
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16.7
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26.5

◀ ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ ▶

In [20]: ▶| `data.head(10)`

Out[20]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	38.0	1	0	PC 17599	71
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	35.0	1	0	113803	53
6	7	0	1	McCarthy, Mr. Timothy J	male	54.0	0	0	17463	51
10	11	1	3	Sandstrom, Miss. Marguerite Rut	female	4.0	1	1	PP 9549	16
11	12	1	1	Bonnell, Miss. Elizabeth	female	58.0	0	0	113783	26
21	22	1	2	Beesley, Mr. Lawrence	male	34.0	0	0	248698	13
23	24	1	1	Sloper, Mr. William Thompson	male	28.0	0	0	113788	35
27	28	0	1	Fortune, Mr. Charles Alexander	male	19.0	3	2	19950	263
52	53	1	1	Harper, Mrs. Henry Sleeper (Myna Haxtun)	female	49.0	1	0	PC 17572	76
54	55	0	1	Ostby, Mr. Engelhart Cornelius	male	65.0	0	1	113509	61

In [16]: ▶| `data.tail()`

Out[16]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52.
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5.
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83.
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30.
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30.

In [21]: ▶| `data.tail(10)`

Out[21]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	
835	836	1	1	Compton, Miss. Sara Rebecca	female	39.0	1	1	PC 17756	83
853	854	1	1	Lines, Miss. Mary Conover	female	16.0	0	1	PC 17592	39
857	858	1	1	Daly, Mr. Peter Denis	male	51.0	0	0	113055	26
862	863	1	1	Swift, Mrs. Frederick Joel (Margaret Welles Ba...	female	48.0	0	0	17466	25
867	868	0	1	Roebling, Mr. Washington Augustus II	male	31.0	0	0	PC 17590	50
871	872	1	1	Beckwith, Mrs. Richard Leonard (Sallie Monypeny)	female	47.0	1	1	11751	52
872	873	0	1	Carlsson, Mr. Frans Olof	male	33.0	0	0	695	5
879	880	1	1	Potter, Mrs. Thomas Jr (Lily Alexenia Wilson)	female	56.0	0	1	11767	83
887	888	1	1	Graham, Miss. Margaret Edith	female	19.0	0	0	112053	30
889	890	1	1	Behr, Mr. Karl Howell	male	26.0	0	0	111369	30

In [17]: ▶| `data.info()`

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 183 entries, 1 to 889
Data columns (total 12 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  183 non-null    int64
 1   Survived     183 non-null    int64
 2   Pclass       183 non-null    int64
 3   Name         183 non-null    object
 4   Sex          183 non-null    object
 5   Age          183 non-null    float64
 6   SibSp        183 non-null    int64
 7   Parch        183 non-null    int64
 8   Ticket       183 non-null    object
 9   Fare         183 non-null    float64
 10  Cabin        183 non-null    object
 11  Embarked     183 non-null    object
dtypes: float64(2), int64(5), object(5)
memory usage: 18.6+ KB
```

In [18]: ▶| `data.shape`

Out[18]: (183, 12)

In [19]: ▶| `data.describe()`

Out[19]:

	PassengerId	Survived	Pclass	Age	SibSp	Parch	
count	183.000000	183.000000	183.000000	183.000000	183.000000	183.000000	183.00
mean	455.366120	0.672131	1.191257	35.674426	0.464481	0.475410	78.68
std	247.052476	0.470725	0.515187	15.643866	0.644159	0.754617	76.34
min	2.000000	0.000000	1.000000	0.920000	0.000000	0.000000	0.00
25%	263.500000	0.000000	1.000000	24.000000	0.000000	0.000000	29.70
50%	457.000000	1.000000	1.000000	36.000000	0.000000	0.000000	57.00
75%	676.000000	1.000000	1.000000	47.500000	1.000000	1.000000	90.00
max	890.000000	1.000000	3.000000	80.000000	3.000000	4.000000	512.32

In [11]: ▶ | `#check the missing value by record`
            `data.isna()`

Out[11]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
1	False	False	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	False	F
6	False	False	False	False	False	False	False	False	False	False	F
10	False	False	False	False	False	False	False	False	False	False	F
11	False	False	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	...	...	
871	False	False	False	False	False	False	False	False	False	False	F
872	False	False	False	False	False	False	False	False	False	False	F
879	False	False	False	False	False	False	False	False	False	False	F
887	False	False	False	False	False	False	False	False	False	False	F
889	False	False	False	False	False	False	False	False	False	False	F

183 rows × 12 columns

◀ [   ] ▶

In [7]: ▶ | `#check missing value by column`
           `data.isna().any()`

Out[7]:
```
PassengerId    False
Survived       False
Pclass         False
Name           False
Sex            False
Age             True
SibSp          False
Parch          False
Ticket         False
Fare           False
Cabin           True
Embarked        True
dtype: bool
```

In [8]: ▶ | `#check the number of missing value`
           `data.isna().sum()`

Out[8]:
```
PassengerId      0
Survived         0
Pclass           0
Name             0
Sex              0
Age            177
SibSp            0
Parch            0
Ticket           0
Fare             0
Cabin          687
Embarked         2
dtype: int64
```

In [9]: ▶|
```python
data=data.dropna()
```

In [10]: ▶|
```python
#to fill missing value by mean value
data["Age"].fillna(29.699118)
```

Out[10]:
```
1      38.0
3      35.0
6      54.0
10      4.0
11     58.0
       ...
871    47.0
872    33.0
879    56.0
887    19.0
889    26.0
Name: Age, Length: 183, dtype: float64
```

In [12]: ▶|
```python
data.isna()
```

Out[12]:

	PassengerId	Survived	Pclass	Name	Sex	Age	SibSp	Parch	Ticket	Fare	C
1	False	False	False	False	False	False	False	False	False	False	F
3	False	False	False	False	False	False	False	False	False	False	F
6	False	False	False	False	False	False	False	False	False	False	F
10	False	False	False	False	False	False	False	False	False	False	F
11	False	False	False	False	False	False	False	False	False	False	F
...	...	...	...	...	...	...	...	...	...	...	
871	False	False	False	False	False	False	False	False	False	False	F
872	False	False	False	False	False	False	False	False	False	False	F
879	False	False	False	False	False	False	False	False	False	False	F
887	False	False	False	False	False	False	False	False	False	False	F
889	False	False	False	False	False	False	False	False	False	False	F

183 rows × 12 columns

In [13]: ▶|
```python
data.isna().sum()
```

Out[13]:
```
PassengerId    0
Survived       0
Pclass         0
Name           0
Sex            0
Age            0
SibSp          0
Parch          0
Ticket         0
Fare           0
Cabin          0
Embarked       0
dtype: int64
```

In [ ]: ▶|